



# A Fuzzy Similarity Based Method for Signal Reconstruction during Plant Transients

Piero Baraldi, Francesco Di Maio, Davide Genini, Enrico Zio

## ► To cite this version:

Piero Baraldi, Francesco Di Maio, Davide Genini, Enrico Zio. A Fuzzy Similarity Based Method for Signal Reconstruction during Plant Transients. Prognostics and System Health Management Conference, Oct 2013, Italy. pp.1-6. hal-00926901

**HAL Id: hal-00926901**

**<https://hal.science/hal-00926901>**

Submitted on 10 Jan 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A Fuzzy Similarity Based Method for Signal Reconstruction during Plant Transients

Piero Baraldi<sup>a\*</sup>, Francesco Di Maio<sup>a</sup>, Davide Genini<sup>a</sup>, Enrico Zio<sup>a,b</sup>

<sup>a</sup>Energy Department, Politecnico di Milano, Via Ponzio 34/3, 20133 Milano, Italy

<sup>b</sup>Chair on System Science and the Energetic Challenge, European Foundation for New Energy – Paris and Supélec, Paris, France  
piero.baraldi@polimi.it

We consider the problem of missing data in the context of on-line condition monitoring of industrial components by empirical, data-driven models. We propose a novel method for missing data reconstruction based on three main steps: (1) computing a fuzzy similarity measure between a segment of the time series containing the missing data and segments of reference time series; (2) assigning a weight to each reference segment; (3) reconstructing the missing values as a weighted average of the reference segments. The performance of the proposed method is verified on a real industrial application regarding shut-down transients of a Nuclear Power Plant (NPP) turbine.

## 1. Introduction

In this work, we consider the problem of missing data in the context of on-line condition monitoring of industrial components by empirical, data-driven models. On-line condition monitoring aims at informing on the health state of industrial components. When using empirical models, the condition monitoring performance is highly dependent on the availability and quality of the measurements used to establish (train) the model (Baraldi et al., 2009), (Baraldi et al./2011), (Baraldi et al./2012), (Boechat et al./2012), (Reifman/1997), (Zio et al./2012). If, for example, a sensor fails to provide an input value, the condition monitoring model may not be capable of inferring the health state of the component. Therefore, it is important to restore the missing sensor readings to provide a set of complete input data to the condition monitoring model, for its training and during its use (Coble et al./2012), (Seibert et al./2007).

To this purpose, we address the problem of missing data in multidimensional time series, e.g. process signals monitored during turbine start-up transients in nuclear power plants. We assume to have available several examples of the time series, e.g. collection of the signal values measured during several, different turbine start-up transients. A difficulty comes from the fact that the reconstruction of a datum missing at a given time does not depend only from the values of the other signals at that time, but also from previous values on the time series.

In this paper, we present a missing data reconstruction method that we have developed based on a Fuzzy Similarity (FS) method (Zio et al./2010a). A measure of similarity is computed between a set of reference multidimensional time-series segments of given length and the multidimensional segment containing the missing data; then, the missing values are reconstructed as average of the reference segments weighted by their similarity with the segment containing the missing data.

The method is tested on a case study concerning 27 signals measured during shut-down transients of nuclear power plants (NPP) steam turbines. The performance of the proposed method is compared with that of an AAKR method of literature (Baraldi et al./2010).

The remainder of the paper is organized as follows: Section 2 introduces the problem that we want to address; Section 3 describes the proposed FS-based method; Sections 4 illustrates the application of the method to an industrial case study; finally, Section 5 draws the conclusion of the work.

## 2. Problem Statement

We consider a training dataset containing  $M$  different  $J$ -dimensional realizations of a time series, hereafter called trajectories, and indicated by  $\bar{X}_m^{tr}$ ,  $m=1, \dots, M$ . These reference trajectories are all complete, i.e. they do not suffer of any missing data. For simplicity of illustration, the reference trajectories are assumed all to have same time length  $T$ . The generic element  $x_m^{tr}(k, j)$  of  $\bar{X}_m^{tr}$  indicates the value of signal  $j$  of trajectory  $m$  at time  $k$ .

The objective of this work is the reconstruction of missing data in a (test) trajectory,  $\bar{\bar{X}}$ , that we are measuring. The length of  $\bar{\bar{X}}$  can be shorter than  $T$ . We consider that the values of only one signal, hereafter referred to as  $j_{miss}$ , are missing in a single time window from time  $t - \varphi$  until the present time  $t$ . The generic element of the test trajectory,  $x(k, j)$ , indicates the value of signal  $j$  at time  $k \leq t$ . The obtained reconstruction of a missing datum of signal  $j_{miss}$  will be indicated by  $\hat{x}(k, j_{miss})$ ,  $t - \varphi \leq k \leq t$ .

### 3. The Fuzzy Similarity-based reconstruction method

The proposed method for missing data reconstruction is based on four main steps:

- (1) compute the point-wise difference between the segment of the test trajectory containing the missing data and the segments obtained from the reference trajectories. At the present time  $t$ , the segment of test trajectory containing only the most recent  $L_t$  measurements of signal  $j$  is denoted as  $\bar{x}_t(j) = [x(t - L_t + 1, j); x(t - L_t + 2, j), \dots, x(t, j)]$  and the generic segment of length  $L_t$  of signal  $j$  in reference trajectory  $m$  which ends at time  $k$  is denoted as  $\bar{x}_m^{tr}(k, j) = \bar{x}_m^{tr}(k - L_t + 1 : k, j)$ . The point-wise difference is computed taking into account all the available  $L_t$  measurements for the signals  $j \neq j_{miss}$  without missing data in the test trajectory and only the available measurements for signal  $j_{miss}$ .

The point-wise difference  $\delta_m^2(k, j)$  between the  $L_t$  elements of the  $k$ -th segment of the  $m$ -th reference trajectory  $\bar{x}_m^{tr}(j)$  and the elements of the test time segment  $\bar{x}_t(j)$  of the  $j$ -th signal, is given by:

$$\delta_m^2(k, j) = \left| \bar{x}_m^{tr}(j) - \bar{x}_t(j) \right|^2 \quad (1)$$

The distance is finally reduced to [Zio et al., 2010b]:

$$\delta_m^2(k) = \frac{\sum_{j=1}^J \delta_m^2(k, j)}{J} \quad (2)$$

- (2) compute a measure of fuzzy similarity  $\mu_m$  between multidimensional segments of the  $M$  reference trajectories in a time window of length  $L_t$  and the most recent segment of length  $L_t$  of the test trajectory containing the missing datum (Zio et al./2010a), (Zio et al./2010b), (Zio et al./2010c). To account for a gradual transition between 'similar' and 'non-similar' we introduce an "approximately zero" fuzzy set taken, in this work, as a bell-shaped function:

$$\mu_m(k) = e^{-\left( \frac{\ln(\alpha)}{\beta^2} \delta_m^2(k) \right)} \quad (3)$$

The parameters  $\alpha$  and  $\beta$  are set by the analyst: the larger the value of the ratio  $\ln(\alpha)/\beta^2$ , the narrower the fuzzy set and the stronger the definition of similarity [Zio et al., 2010a].

- (3) assign a weight  $w_m$  to each reference segment; the weight is chosen proportional to the similarity of the reference segments to the test trajectory:

$$w_m = \mu_m e^{-\frac{1}{\beta}(1 - \mu_m(k))} \quad m = 1, \dots, M \quad (4)$$

- (4) reconstruct the missing datum as a weighted average of the reference segments:

$$\hat{x}(t, j_{miss}) = \frac{\sum_{m=1}^M \sum_{k=L_t}^T w_m(k) x_m^{tr}(k, j_{miss})}{\sum_{m=1}^M \sum_{k=L_t}^T w_m(k)} \quad (5)$$

where  $x_m^{tr}(k, j_{miss})$  is the  $k$ -th segment of the  $m$ -th reference trajectory.

#### 4. Application to an industrial case study

The industrial case study concerns the operation of nuclear power plant turbines during shut-down transients. We consider the values of  $J = 27$  signals taken at  $T=4500$  time steps in 148 different transients. Most of the signals refer to temperatures measured in different parts of the turbines (Baraldi et al., 2010). Figure 1 shows some examples of signal evolutions during different plant transients.

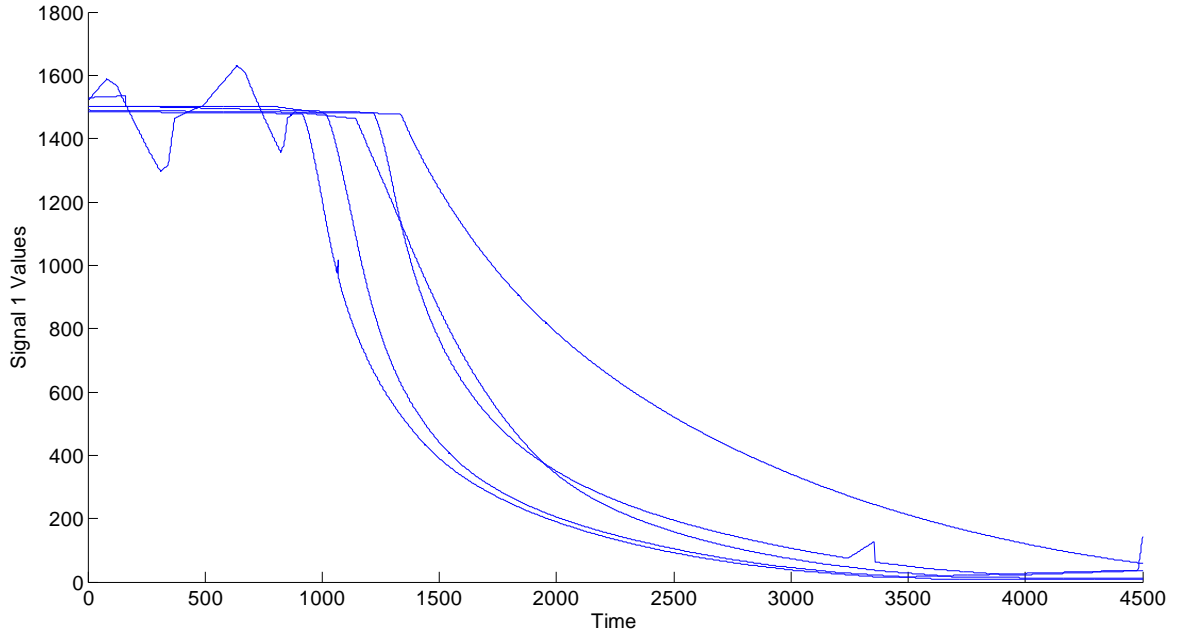


Figure 1: Evolution of a signal in different plant transients.

The performance of the proposed model for missing data reconstruction is here evaluated in terms of its accuracy, i.e. the ability of providing correct reconstructions of missing data. The metric used is the Mean Square Error (MSE) between the reconstructions provided by the model and the true values (Baraldi et al./2010), averaged over a number of different test trajectories. In order to compute the MSE metric, we perform the reconstruction of signal segments whose true values is known, but for which we assume to have missing data in time intervals of  $\varphi=20$  time instants in a single signal.

The values of the FS-method parameters  $\alpha$  and  $\beta$ , and of the length of the time segment have been set according to a trial and error procedure (Table 1).

Figure 2 shows the overall accuracy of the FS-based and of the AAKR methods in the reconstruction of different plant signals in 27 test transients, according to a leave-one-out procedure (Baraldi et al./2012), (Polikar/2007).

Table 1: Optimal values of the parameters  $\alpha$  and  $\beta$  and the length of the time segment  $L_t$ .

Parameter	value
$L_t$	10
$\alpha$	0.2
$\beta$	0.5

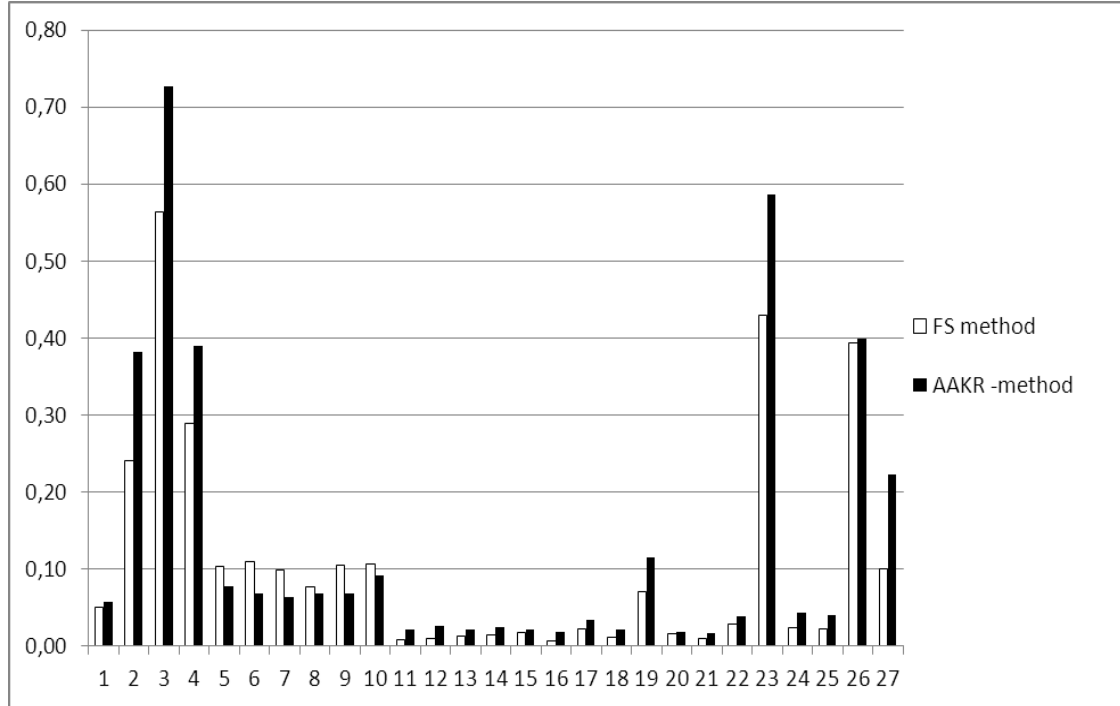


Figure 2: Average accuracy in the reconstruction of the 27 signals.

The performance of the FS-based reconstruction method is more satisfactory than that of the AAKR method. In particular, Figure 3 compares the FS-based method (dotted line with circles) and the AAKR (dashed line) reconstructions of some signals in a transient, assuming missing data on a time window from instant  $t_A=81$  to instant  $t_B=100$ . Notice that for several signals such as  $j=2, 14, 18, 22$  and  $27$  the AAKR method provides reconstructions which remarkably deviate from the true signal values, whereas the reconstructions of the FS-based method are more accurate.

With respect to signal 26, both methods provide very inaccurate reconstructions. This is due to the fact that the behaviour of this transient in signal 26 is very different from the behaviour of the signal in all the considered reference trajectories. Since the reconstruction is a weighted mean of the training values, this leads to the inaccurate signal reconstruction.

## Conclusions

A fuzzy similarity-based method for missing data reconstruction has been proposed in the context of on-line condition monitoring of industrial components. The method allows performing signal reconstructions in multidimensional time series. It has been applied with success to a real industrial application concerning the reconstruction of missing data in nuclear component transients, and it has been shown superior to an AAKR-based method of literature.

Given the difficulty of the signal reconstruction task in situations characterized by the presence of long segments containing missing values, we think that it would be important to associate the signal reconstruction with an estimate of its degree of confidence, which should take into account the amount and quality of the information used to perform the reconstruction. This will be object of future research activity. Indeed, future work will be devoted to estimate the degree of confidence in the provided signal reconstruction, taking into account the amount of information available in the reference trajectories used to perform the reconstruction.

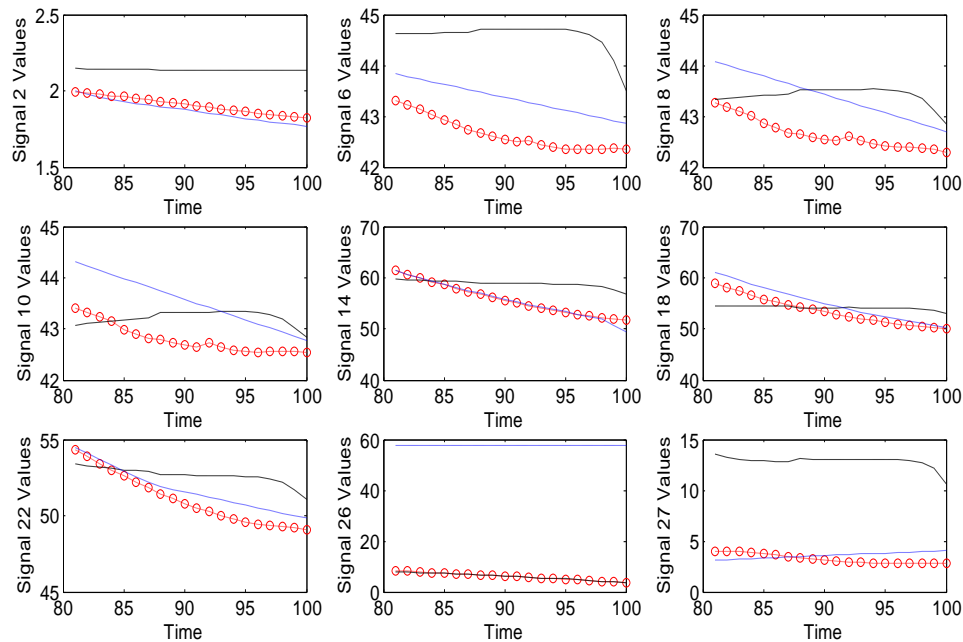


Figure 3: Reconstruction of a time window of 20 time instants in some signals. The true value is represented by a dotted line, the FS-based reconstruction by a continuous line and the AAKR reconstruction by a dashed line.

## Acknowledgments

The authors are thankful to EDF R&D STEP Department for providing the data for the case study.

## References

- Baraldi P., Zio E., Gola G., Roverso D., Hoffmann M., 2009, A procedure for the reconstruction of faulty signals by means of an ensemble of regression models based on principal components, Nuclear Plant Instrumentation, Controls, and Human Machine Interface Technology.
- Baraldi P., Canesi R., Zio E., Seraoui R., Chevalier R., 2010, Signal Grouping for Condition Monitoring of Nuclear Power Plant Components, Seventh American Nuclear Society International Topical Meeting on Nuclear Plant Instrumentation, Control and Human Machine Interface Technology, NPIC&HMIT 2010, Las Vegas, Nevada, 7-11 November.
- Baraldi P., Razavi-Far R., Zio E., 2011, Bagged Ensemble of FCM Classifiers for Nuclear Transient Identification, Annals of Nuclear Energy; Vol. 38 (5), pp 1161-1171.
- Baraldi P., Di Maio F., Pappaglione L., Zio E., Seraoui R., 2012, Condition Monitoring of Electrical Power Plant Components During Operational Transients, Proceedings of the Institution of Mechanical Engineers, Part O, Journal of Risk and Reliability, 226(6) 568–583, 2012.
- Baraldi P., Di Maio F., Zio E., Saucó S., Droguett E., Magno C., 2012, Sensitivity analysis of scale deposition on equipment of oil wells plants, Chemical Engineering Transactions, 26, Chemical Engineering Transactions, 26, pp. 327-332.
- Boechat A.A., Moreno U.F., Haramura D., 2012, On-line calibration monitoring system based on data-driven model for oil well sensors, Proceedings of the 2012 IFAC Workshop on Automatic Control in Offshore Oil and Gas Production.
- Coble J.B., Ramualli P., Bondi J., Hines J.M., Upadhyaya, B.R., 2012, Prognostics and Health Management in Nuclear Power Plants: A Review of Technologies and Applications, National Technical Information Service.

- Polikar, R., 2007, Bootstrap-inspired techniques in computational intelligence, *IEEE Signal Processing Magazine* 59, 59–72.
- Reifman J., 1997, Survey of artificial intelligence methods for detection and identification of component faults in nuclear power plants, *Nuclear Technology*, 119, pp. 76-97.
- Seibert R., Ramsey C.R., Garvey D.R., Hines J.W., Robison B.H., Outten S.S., 2007, Verification of helical tomotherapy delivery using autoassociative kernel regression, *Medical Physics*, Vol. 34.
- Zio E., Di Maio F., 2010a, A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system, *Reliability Engineering and System Safety*, Vol. 95.
- Zio E., Di Maio F., Stasi M., 2010b, A Data Driven Approach for Predicting Failure Scenarios in Nuclear Systems, *Annual of Nuclear Energy*, 37, 482-491.
- Zio E., Di Maio F., A fuzzy similarity-based Method for Failure Detection and Recovery Time Estimation, *International Journal of Performability Engineering*, Vol. 6, No. 5, 2010.
- Zio E., Broggi M., Golea L.R., Pedroni N., 2012, Failure and reliability predictions by infinite impulse response locally recurrent neural networks, *Chemical Engineering Transactions*, 26, pp. 117-122.